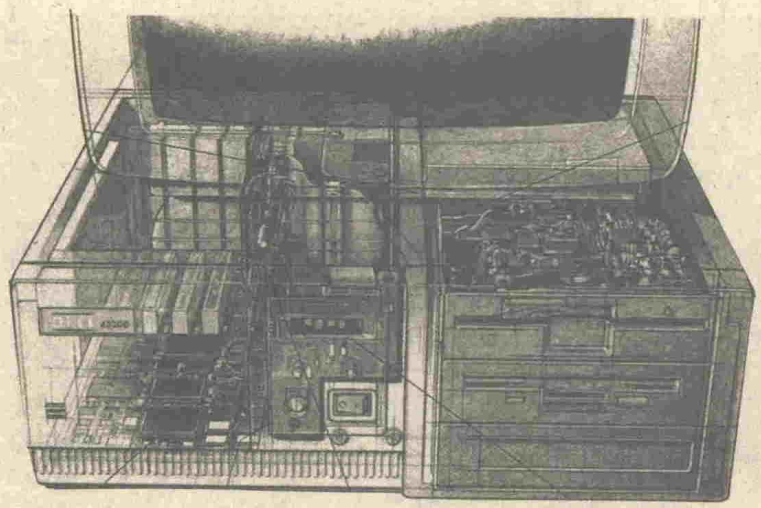
ชัยยงค์ วงศ์ชัยสุวัฒน์

เพิ่มความจุของดิสค์ ด้วยการบีบอัดข้อมูล (1)



เพิ่มความจุของฮาร์ดดิสค์ด้วยการปรับเปลี่ยนโครงสร้างของดิสค์ให้แตกต่างกับการจัดการของดอสปกติ

ในสมัยก่อนเวลาเราพูดถึงฮาร์ดดิสค์ 20 หรือ 30 เมกะไบต์ เรามักคิดว่ามันมากเกินกว่าการนำไปใช้งานจริง ในปัจจุบันฮาร์ดดิสค์ขนาดความจุ 30 เมกะไบต์ ถือว่าเล็กเกินไปเสียแล้ว เพราะซอฟต์แวร์สมัยใหม่จำเป็นต้องใช้เนื้อที่ดิสค์สำหรับเก็บไฟล์โปรแกรมและไฟล์ของระบบต่าง ๆ หลาย ๆ เมกะไบต์ มีผู้กล่าวไว้ว่า ในสภาพของการใช้งานไมโครคอมพิวเตอร์ปัจจุบัน ควรมีฮาร์ดดิสค์ที่มีความจุมากกว่า 100 เมกะไบต์ จึงจะเหมาะสม โดยราคาของฮาร์ดดิสค์ก็ได้ลดต่ำลงเป็นไปได้มากทีเดียว

เมื่อเรามีระบบไมโครคอมพิวเตอร์ ที่มีฮาร์ดดิสค์อยู่เรียบร้อยแล้ว แต่ขนาดความจุของฮาร์ดดิสค์น้อยเกินไป วิธีการแก้ไขก็ต้องจัดซื้อหรือหาฮาร์ดดิสค์ที่มีขนาดความจุสูงขึ้นมาเสริมหรือแทนที่ฮาร์ดดิสค์เดิม การซื้อหามาใหม่นี้ต้องมีค่าใช้จ่ายสูง ซึ่งเราอาจมีงบไม่เพียงพอ ปัญหาที่ขนาดความจุของฮาร์ดดิสค์มีไม่เพียงพอนี้ได้เริ่มบรรเทาลงบางแล้ว เพราะในท้องตลาดเริ่มมีซอฟต์แวร์ยูทิลิตี้มาช่วยเพิ่มความจุของดิสค์ให้สูงขึ้นอีกราวเท่าตัว โดยใช้วิธีการบีบอัดข้อมูลให้เล็กลงก่อนนำไปเก็บในดิสค์ และเมื่อเวลาต้องเรียกใช้ไฟล์นั้น ก็จะทำการแผ่ขยายข้อมูลกลับนำสู่สถานภาพเดิมทุกประการ ข้อดีของซอฟต์แวร์นี้ทำให้เราลดการลงทุนซื้อฮาร์ดดิสค์ใหม่ได้ดี แต่จะมีข้อเสียที่การทำงานอาจช้ากว่าการติดต่อกับดิสค์โดยตรง เพราะจะต้องนำผ่านการบีบอัดและแผ่ขยายข้อมูลไฟล์ทุกครั้ง

วิธีการบีบอัดและแผ่ขยายข้อมูลไฟล์นี้ อาจอยู่ในรูปของฮาร์ดแวร์ที่ทำเป็นการ์ดมาเสียบเพิ่ม หรืออยู่ในรูปของซอฟต์แวร์ หรืออยู่ในรูปผสมที่มีทั้งฮาร์ดแวร์และซอฟต์แวร์รวมกัน เนื่องจากซอฟต์แวร์เพื่อการเพิ่มขนาดความจุของดิสค์เป็นสองเท่า เริ่มได้รับความนิยมนำมาใช้กันมากขึ้นในปัจจุบัน มุมโอเอสฉบับนี้จึงขอนำเสนอเรื่องราว หรือวิธีการของซอฟต์แวร์กลุ่มนี้ โดยจะยกตัวอย่างซอฟต์แวร์ Stacker 2.0 ที่วารสาร PC Magazine ฉบับวันที่ 28 มกราคม 2535 จัดให้เป็นซอฟต์แวร์เพิ่มความจุดิสค์ที่ดีที่สุด เชื่อถือได้มากที่สุด และโปรแกรม SuperStor ของบริษัท AddStor Inc. ซึ่งซอฟต์แวร์ DR DOS 6.0 ได้รวมโปรแกรมหลังนี้เข้าไปรวมอยู่ด้วย คิดว่าจะเป็นประโยชน์ต่อผู้อ่านอีกเช่นเคย

ประเภทของซอฟต์แวร์บีบอัดข้อมูล

ในลักษณะของการใช้งานซอฟต์แวร์เพื่อการบีบอัดข้อมูล อาจแบ่งได้เป็นสองจำพวกคือ batch compression และ on-the-fly compression ประเภทแรก จะเป็นยูทิลิตี้ที่เรานิยมกันมากในปัจจุบัน ซอฟต์แวร์ที่เราเพิ่งซื้อมาใหม่บนแผ่นดิสค์

เพียงสองสามแผ่น เมื่อติดตั้งเข้าในฮาร์ดดิสค์จะใช้พื้นที่ถึงหลาย ๆ เมกะไบต์ ก็เพราะซอฟต์แวร์เหล่านั้นมีการบีบอัดข้อมูลให้เล็กลง เพื่อเก็บในแผ่นดิสค์จำนวนน้อยแผ่น การติดตั้งซอฟต์แวร์จะเป็นการเรียกไปยังโปรแกรมแผ่ขยาย เพื่อยืดซอฟต์แวร์ไปสู่ขนาดเดิม ยูทิลิตี้ในกลุ่มนี้ที่โด่งดังและนิยมใช้กันมาก ได้แก่ ARC ของบริษัท System Enhancement Associates ที่มีมาก่อนใคร ๆ ทั้งหมด ต่อมาจึงมี LHARC จากประเทศญี่ปุ่น และ PKZIP ของบริษัท PKWARE ซึ่งบริษัทหลังสุดนี้ยังมีผลิตภัณฑ์อื่น ๆ ตามมาอีกหลายตัว แต่ละตัวก็มีขีดความสามารถสูงขึ้น (ผู้อ่านที่สนใจอาจหาอ่านจากวารสารไมโครฯ ฉบับเก่า ๆ ที่เคยนำมาเสนอแล้ว)

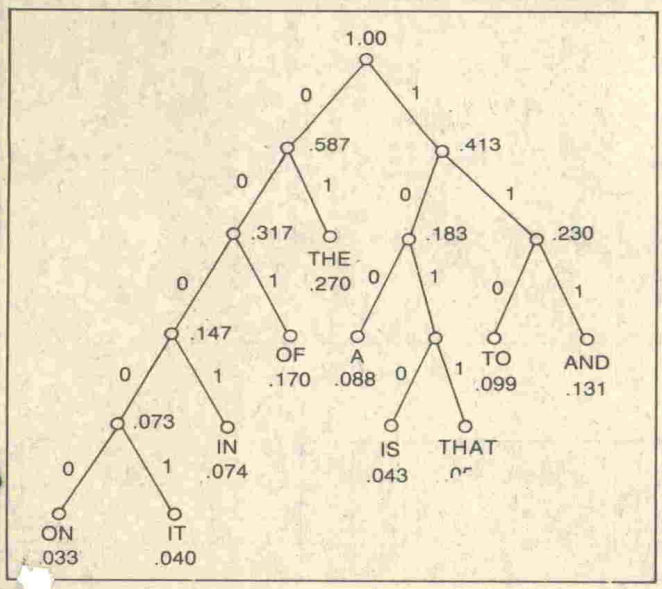
ประเภทหลังที่เรียก On-the-fly หรือ Real-time Compression จะเป็นโปรแกรมที่ทำงานอยู่หลังฉาก ที่จะทำการบีบอัดและแผ่ขยายไฟล์ ขณะที่มีการเรียกติดต่อ คือจะมีการนำไปเขียนบันทึก เก็บในดิสค์แบบบีบอัดข้อมูล และถ้าจะมีการเรียกมาใช้ก็จะนำข้อมูลไฟล์จากในดิสค์มาทำการแผ่ขยายให้กลับสู่ส่วนสภาพเดิมได้อย่างถูกต้องไม่ผิดพลาด ซึ่งการทำงานนี้เป็นการทำภายในไม่กระทบกระเทือนต่อระบบดอสและผู้ใช้งาน ผู้ใช้งานเองจะไม่รู้เลยว่ามีการบีบอัดหรือแผ่ขยายข้อมูลอยู่ มุมโอเอสจะกล่าวถึงการบีบอัดและแผ่ขยายไฟล์แบบ on-the-fly หรือ real-time นี้ เพราะมีแนวโน้มของการนำไปใช้มากขึ้น

เทคนิคการบีบอัดข้อมูล

เทคนิคการบีบข้อมูลอาจแบ่งได้เป็นสองพวกคือ พวกที่อาศัยข้อมูลทางสถิติและดูความน่าจะเป็นของตัวอักษรต่าง ๆ (statistical) อีกพวกหนึ่งอาศัยหลักการของพจนานุกรม (Dictionary) โดยที่กลุ่มของตัวอักษรที่เรียงต่อกัน อาจเกิดซ้ำ ๆ ได้บ่อย ภายในเอกสารชุดหนึ่ง ๆ ดังนั้นวิธีพจนานุกรม จะเป็นการนำกลุ่มคำมาเข้ารหัส สร้างเป็นดัชนีของการชี้ค้นหาคำจริงในพจนานุกรม ซึ่งวิธีนี้จะช่วยลดจำนวนตัวอักษรและประหยัดพื้นที่ได้มาก

รูปที่ 1 เป็นตัวอย่างของการบีบอัดข้อมูลด้วยวิธีทางสถิติที่เรียกรหัส Huffman โดยคำหรือตัวอักษรแต่ละหน่วยจะมีค่าความน่าจะเป็นจากการศึกษาวิจัยมาแล้ว ในรูปที่ 1 ก. เป็นต้นไม้ของรหัส Huffman ที่มียอดเป็นค่าความน่าจะเป็นได้เท่ากับหนึ่ง สำหรับคำภาษาอังกฤษที่นิยมใช้กันมากในภาษาอังกฤษ การแตกกระจายของต้นไม้ในสองทางสำหรับรหัสทางค่าศูนย์และทางค่าหนึ่ง ทำให้เราได้รหัสของค่าต่าง ๆ ดังรูปที่ 1 ข. ทำให้ค่าความยาวของรหัสมีจำนวนบิตไม่เท่ากัน แปรค่าตามโอกาสความถี่ของการเกิดคำเหล่านั้น ค่าที่พบบ่อย ๆ จะได้จำนวนบิตที่น้อยกว่า

รูปที่ 2 เป็นตัวอย่างการบีบอัดข้อมูลที่จัดเป็นแบบวิธีพจนานุกรม โดยรหัสตัวอักษรของระบบ Braille ที่ใช้กับ



รูปที่ 1 ก.

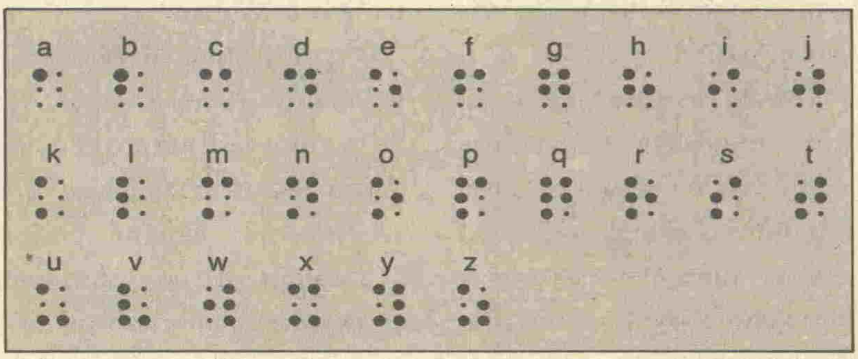
การเข้ารหัส	ความน่าจะเกิด	ค่ารหัส	ความยาวรหัส
the	.270	01	2
of	.170	001	3
and	.131	111	3
to	.099	110	3
a	.088	100	3
in	.074	0001	4
that	.052	1011	4
is	.043	1010	4
it	.040	00001	5
on	.033	00000	5

รูปที่ 1 ข.

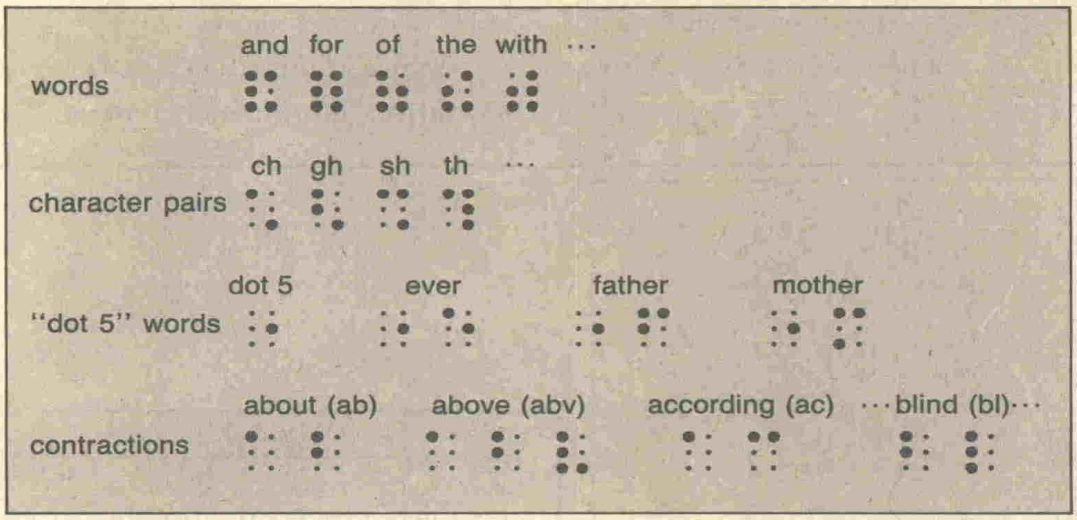
รูปที่ 1 ตัวอย่างรหัส **Huffman**

คนตาบอด ในรูปที่ 2 ก. มีการเข้ารหัสเป็นคำและกลุ่มตัวอักษร รูปที่ 2 ข. ดังนั้นข้อความ foundation for the blind สามารถจัดเป็นรหัสดังรูปที่ 2 ค. ซึ่งจะช่วยลดจำนวนชุดของจุดที่ใช้แทนคำและตัวอักษรให้น้อยลง

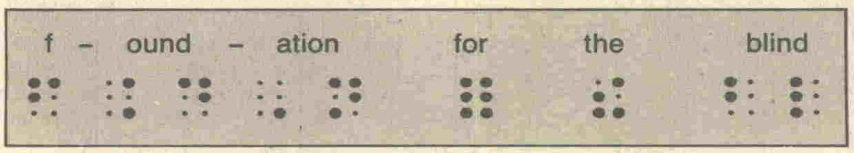
การบีบอัดข้อมูลด้วยวิธีพจนานุกรม ยังสามารถแบ่งแยกย่อยได้เป็น 3 ชนิดด้วยกันคือ ชนิด Static ชนิด Semiadaptive และชนิด อะแดปตีฟ (adaptive) รูปที่ 3 แสดงขั้นตอนการบีบอัดและแผ่ขยายข้อมูล ซึ่งถ้าเป็นระบบที่เชื่อถือได้ ข้อมูลต้นทางและข้อมูลปลายทางจะต้องเหมือนกันทุกประการ ชนิดของการบีบอัดข้อมูลด้วยวิธีพจนานุกรม จะขึ้นกับรูปแบบที่ใช้ในการบีบอัดและแผ่ขยายข้อมูล หากการเข้ารหัสและการถอดรหัส มีการใช้รูปแบบการบีบอัดและการแผ่ขยายแบบคงที่ ไม่เปลี่ยนแปลงตามข้อความหรือข้อมูลที่เข้ารหัส เราจะเรียกว่า การเข้ารหัสชนิด Static โดยตารางเทียบรหัส หรือ



รูปที่ 2 ก.

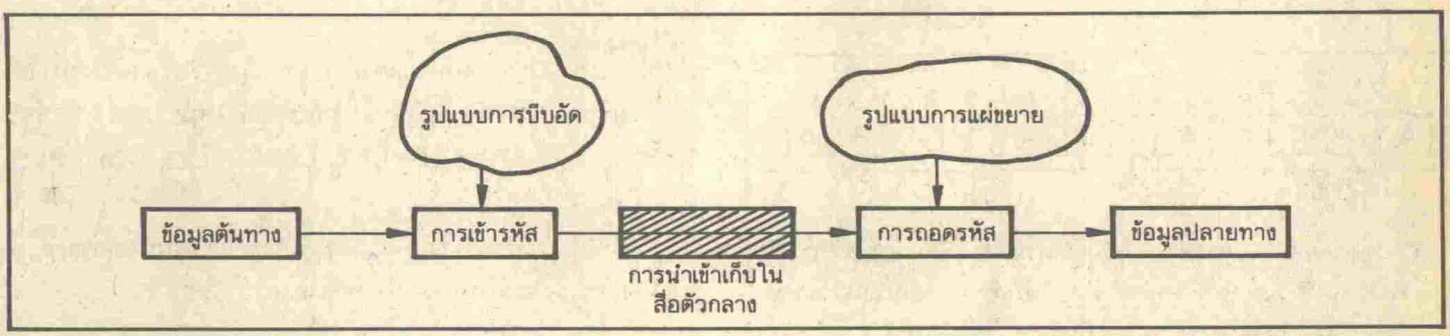


รูปที่ 2 ข.

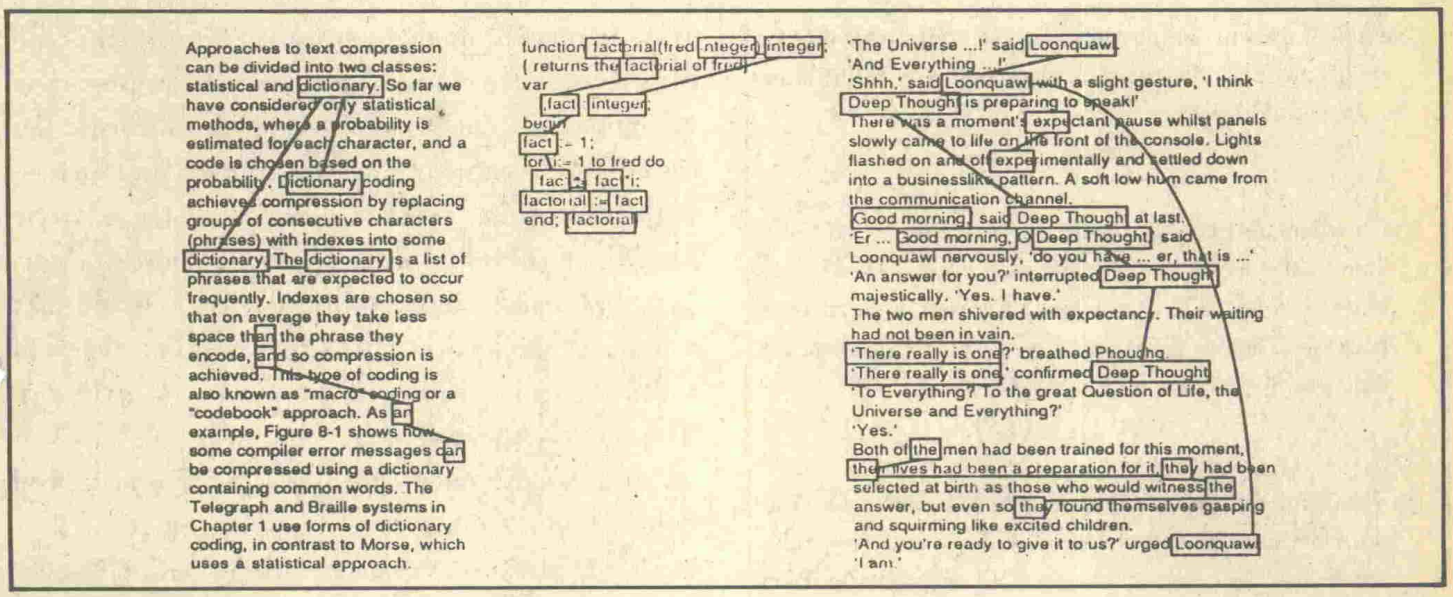


รูปที่ 2 ค.

รูปที่ 2 ระบบ Braille ที่ใช้การบีบอัดข้อมูลแบบพจนานุกรม



รูปที่ 3 การบีบอัดและแผ่ขยายข้อมูล



รูปที่ 4 หลักการเข้ารหัส Ziv Lempel แบบพจนานุกรม

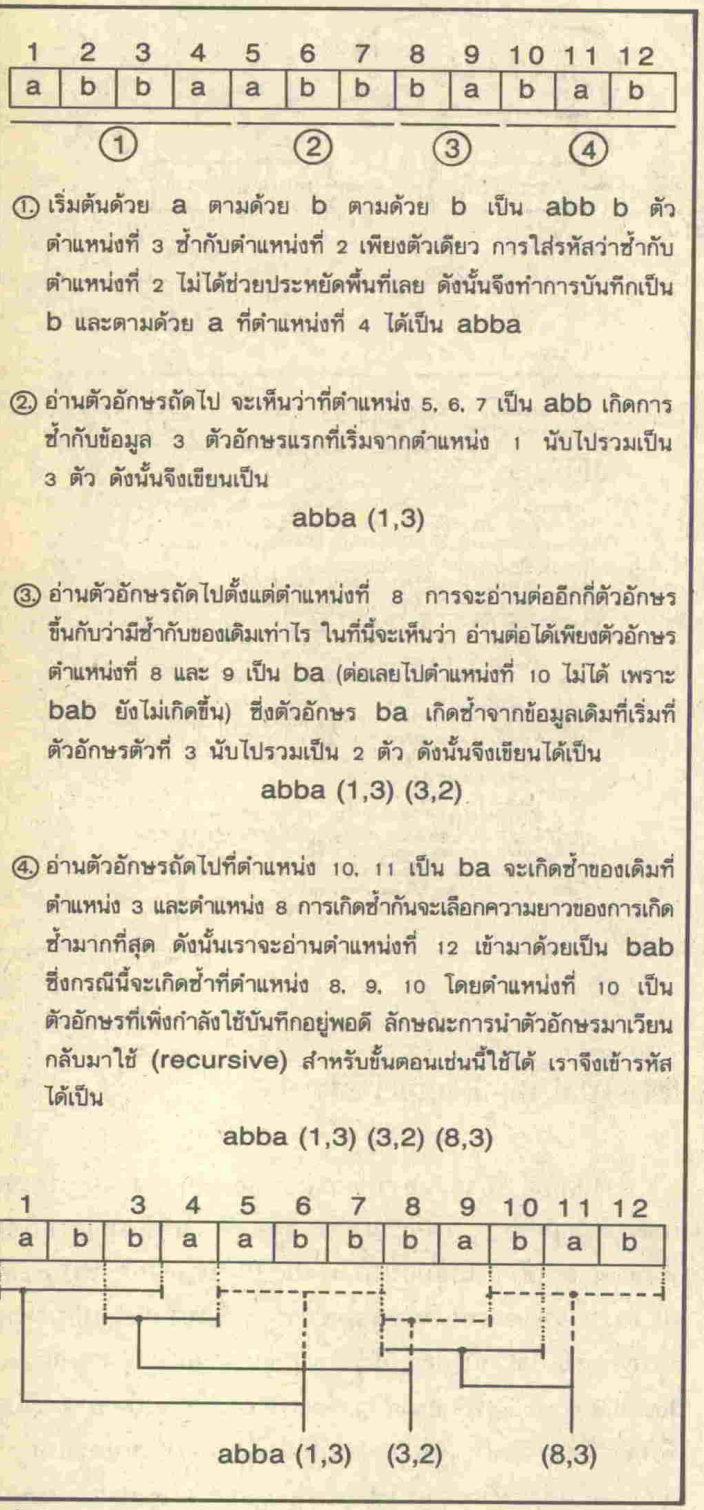
พจนานุกรมของคำหรือกลุ่มตัวอักษรจะต้องคงที่ตายตัวทั้งตอนเข้ารหัสและการถอดรหัส

วิธีแบบ Static ข้างต้นไม่มีมากนัก เพราะถ้าข้อความหรือข้อมูลต่างสาขามากมาย ขนาดของตารางเทียบรหัสจะยิ่งใหญ่โตมากตามไปด้วย หนทางแก้ไขทางหนึ่งก็คือ มีการบอกรูปแบบของการแผ่ขยายข้อมูลมาใช้ตารางเทียบรหัสเดียวกันที่เล็กกว่า วิธีนี้เรียกว่า Semiadaptive ข้อเสียของแบบ Semiadaptive จะทำให้ต้องเสียเวลาอ่านและตรวจดูว่าข้อความหรือข้อมูลเป็นสาขาใดก่อน ขณะเดียวกันอาจต้องมีการเก็บตารางเทียบรหัสไว้ภายในด้วย

วิธีสุดท้ายเรียก Adaptive ซึ่งการเข้ารหัสและการถอดรหัสจะมีการใช้รูปแบบที่มีการสร้างตารางเทียบรหัสจากข้อความหรือข้อมูลเดิมที่ผ่านมาแล้ว วิธีนี้จะเป็นแบบที่เรานำมาใช้ในการบีบอัดและแผ่ขยายเพิ่มความจุของดิสค์เป็นสองเท่า ซึ่งรายละเอียดจะได้กล่าวในหัวข้อถัดไป

ขั้นตอนวิธี Lempel-Ziv

ผลิตภัณฑ์ทั้งด้านฮาร์ดแวร์และซอฟต์แวร์ของการบีบอัดและแผ่ขยายข้อมูล เพื่อเพิ่มความจุของดิสค์เป็นสองเท่าเกือบทั้งหมด จะใช้วิธีบีบอัดแบบ on-the-fly หรือเรียลไทม์ (real-time) ซึ่งอาศัยเทคนิคของพจนานุกรมแบบอะแดปตีฟตามที่กล่าวมาแล้ว ขั้นตอนวิธีที่นิยมใช้มาจากแนวความคิดของ Jacob Ziv และ Abraham Lempel ที่ให้ไว้ตั้งแต่ปี พ.ศ. 2520 ที่กล่าวถึง การเข้ารหัสด้วยเทคนิคพจนานุกรมแบบอะแดปตีฟว่า การเข้ารหัสส่วนของอินพุตในอนาคตสามารถคัดลอกจากบัฟเฟอร์ที่ประกอบด้วยข้อมูลเอาต์พุตที่ผ่านไปแล้ว ซึ่งแนวความคิดนี้หมายถึงว่า คำหรือวลีที่เคยมีมากแล้ว อาจเกิดซ้ำ ๆ ถ้าเรานำคำหรือวลีนั้นเก็บเข้ารหัสไว้เหมือนพจนานุกรมดังรูปที่ 4 ที่วลีจะเข้ารหัสชี้ไปยังวลีที่เคยมีมาก่อน



รูปที่ 5 ตัวอย่างรหัส Ziv-Lempel

วิธีการของ Ziv และ Lemple จะรวมเรียกว่า การเข้ารหัส LZ คือเป็นต้นฉบับของการบีบอัดข้อมูล ซึ่งมีผู้นำไปปรับปรุงเพิ่มขยายกันอีกมากมาย และเรียกชื่อแตกต่างกันไป แต่ส่วนใหญ่ยังคงตัวอักษร 2 ตัวแรกไว้ เช่น LZR, LZSS, LZB, LZH, LZW, LZC, LZT, LZMW, LZJ หรือ LZFG เป็นต้น

ตัวอย่างการเข้ารหัส Ziv Lempel อาจแสดงดังรูปที่ 5 ของการเข้ารหัสข้อมูล abbaabbbabab

ผลิตภัณฑ์บีบอัดและแผ่ขยายข้อมูล

ผลิตภัณฑ์บีบอัดและแผ่ขยายข้อมูลแบบเรียลไทม์ เพื่อเพิ่มขนาดความจุฮาร์ดดิสค์เป็นสองเท่าที่ผลิตออกจำหน่ายจะใช้ขั้นตอนวิธี Lempel-Ziv เป็นหลักสำคัญ แต่มีการเพิ่มเติมเทคนิคอื่นเสริม เช่น เทคนิคที่พยายามคงความถูกต้องของข้อมูลไว้ครบทุกบิต ที่เรียกว่า lossless technique เพราะข้อมูลที่บีบอัดให้ลดพื้นที่เก็บข้อมูลแล้ว การเรียกข้อมูลคืนมาในรูปปกติจะต้องเหมือนของเดิมทุกประการ ไม่ให้มีอะไรขาดหายไปโดยเด็ดขาด บริษัทผู้ผลิตจะนำขั้นตอนวิธีเข้าไปรวมเป็นฮาร์ดแวร์ภายใน ตัวประมวลผลร่วมเป็นการ์ดเสียบ หรือเข้าไปรวมเป็นซอฟต์แวร์

ผลิตภัณฑ์บีบอัดและแผ่ขยายข้อมูล เพิ่มความจุดิสค์ที่เด่น ๆ ได้แก่

- DoubleDisk, Version 2.5 บริษัท Verisoft Systems
- Expanz! Plus ของบริษัท InfoChip Systems
- Stacker, Version 2.0 ของบริษัท Stac Electronics
- SuperStor, Version 1.3 ของบริษัท AddStor

ผลิตภัณฑ์ที่เป็นซอฟต์แวร์คือ Double Disk 2.5, Stacker 2.0 และ SuperStor 1.3 ส่วนที่ทำเป็นฮาร์ดแวร์แบบการ์ดเสียบในช่องสล้อตคือ Expanz! Plus เป็นการ์ด coprocessor แบบ 8 บิต ในส่วนของบริษัท Stac Electrenics ผู้ผลิต Stacker นอกจากจะมีซอฟต์แวร์ Stacker 2.0 แล้วยังมีผลิตภัณฑ์เป็นฮาร์ดแวร์แบบการ์ดเสียบอีก 3 แบบ คือ

- Stacker AT/16 coprocessor สำหรับเครื่องไอบีเอ็มตระกูล AT, 286, 386, 486 เครื่องพีเอสทู โมเดล 25, 30, 35, 40 หรือเครื่องเสมอเหมือนอื่น ๆ โดยมีสล้อตเสียบ 16 บิต
- Stacker MC/16 coprocessor สำหรับเครื่องพีเอสทู โมเดล 50 ขึ้นไปที่ใช้บัสแบบ Micro Channel หรือเครื่องเสมอเหมือนอื่น ๆ
- Stacker XT/8 coprocessor สำหรับเครื่องไอบีเอ็มตระกูล XT ที่มีช่องสล้อตเสียบแบบ 8 บิต

ลักษณะพิเศษของผลิตภัณฑ์บีบอัดและขยายข้อมูล

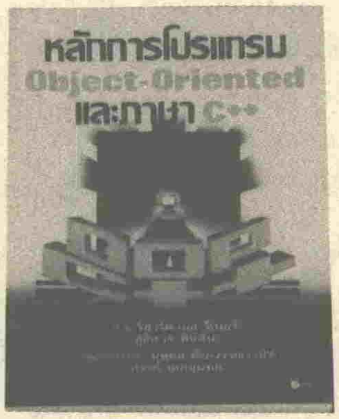
ผลิตภัณฑ์เพิ่มความจุของดิสค์ต่าง ๆ มักจะมีลักษณะพิเศษต่าง ๆ ดังนี้

- ใช้ขั้นตอนวิธีตาม Lempel-Ziv และเพิ่มเสริมเทคนิคเฉพาะตัวของบริษัท
- ขนาดของซอฟต์แวร์ที่เป็นดีไวซ์ไดรเวอร์ ประมาณสี่สิบกว่ากิโลไบต์
- สนับสนุนการใช้หน่วยความจำเพิ่มขยาย EMS
- มียูทิลิตี้เพื่อการนำไฟล์ที่ใช้พื้นที่กระจัดกระจายมารวมติดต่อกัน (defragger)
- มียูทิลิตี้ตรวจสอบและจัดการระบบ
- สามารถเสริมไดรฟ์ที่บีบอัดข้อมูลเพิ่มอีกนับสิบไดรฟ์
- ผลิตภัณฑ์เหล่านี้จะทำงานร่วมกับซอฟต์แวร์ดัง ๆ เช่น ไมโครซอฟต์วินโดวส์ 3.0, นอร์ตันยูทิลิตี้, พีซีทูลส์ และ Super PC-Kwik ได้ดีไม่มีปัญหา
- สามารถบีบอัดและแผ่ขยายข้อมูลกับสื่อหลายชนิด คือ Bernoulli Box แผ่นดิสค์ ฮาร์ดดิสค์แบบ SCSI ฮาร์ดดิสค์แบบ RLL และแรมดิสค์ (นำหน่วยความจำมาจำลองให้เป็นดิสค์)

สรุป

การเพิ่มความจุของฮาร์ดดิสค์ให้สูงขึ้น ด้วยการปรับเปลี่ยนโครงสร้างของดิสค์ให้แตกต่างกับการจัดการของดอสปกติ จะช่วยเพิ่มขีดความสามารถในการเก็บข้อมูลของดิสค์ได้มากขึ้น ทำให้เราสามารถชะลอการจัดหาฮาร์ดดิสค์ที่มีขนาดความจุมาก ๆ มาใช้ได้ระยะหนึ่ง ผลิตภัณฑ์เพื่อการบีบอัดและแผ่ขยายข้อมูลเพื่อเพิ่มความจุกับดิสค์ได้เป็นสองเท่า เริ่มได้รับความสนใจนำมาใช้งานกันมากขึ้น โปรแกรมดอส DR DOS 6.0 ก็รับเอาซอฟต์แวร์ SuperStor ของการเพิ่มความจุดิสค์เป็นสองเท่าเข้ามารวมในโปรแกรมแล้ว ดังนั้นเราจะได้พบเห็นหรือมีโอกาสใช้ซอฟต์แวร์ประเภทมากขึ้น

มุมโอเอสฉบับนี้เป็นการแนะนำซอฟต์แวร์ประเภทนี้ขั้นแรก ในฉบับหน้าจะได้นำเสนอซอฟต์แวร์ Stacker 2.0 ซึ่งถือเป็นซอฟต์แวร์ที่เด่นมากสุด น่าใช้ที่สุดของซอฟต์แวร์กลุ่มนี้ พร้อมกันนี้จะได้นำเสนอโปรแกรม SuperStor ของ DR DOS 6.0 ประกอบด้วย พบกันในฉบับหน้านะครับ



หลักการโปรแกรม
Object-Oriented
และภาษา C++

โดย
ริชาร์ต เอส. วีเนอร์ ลูอิส เจ. ฟินสัน
เรียบเรียงโดย
นพดล ตังควรรณวานิช
สรรค์ เสกขุนทด
■ ราคาเล่มละ 165 บาท

Object-Oriented Programming (OOP) นับว่าเป็นก้าวต่อไปของการพัฒนาระบบซอฟต์แวร์คอมพิวเตอร์ เป็นก้าวที่ห่างไกลจากวิธีการเดิมคือ การเขียนโปรแกรมด้วยภาษาแบบเรียกใช้ฟังก์ชั่น เช่น ภาษา C Pascal, COBOL และภาษาอื่น ๆ เป็นอย่างมาก

หนังสือเล่มนี้ได้เสนอภาษา C++ ซึ่งเป็นภาษา Object-Oriented Programming ที่กำลังได้รับความนิยมอย่างสูง พร้อมได้รวมเอาวิธีการพัฒนาระบบซอฟต์แวร์แบบ OOP เข้าไว้ด้วย และเป็นที่แน่นอนว่าเนื้อหาที่มีทั้งภาษาและวิธีการพัฒนาซอฟต์แวร์ จะทำให้หนังสือเล่มนี้มีความสมบูรณ์แบบอยู่ในตัว

หาซื้อตามร้านหนังสือทั่วไปหรือสั่งซื้อโดยตรงได้ที่ ● **บริษัท ซีเอ็ดยูเคชั่น จำกัด** 800/43-45 ซอยตระกูลสุข ถนนอโศก-ดินแดง ห้วยขวาง กรุงเทพฯ 10400 โดยส่งธนาณัติสั่งจ่าย **ปท.ดินแดง** ในนามบริษัทฯ หรือเช็คส่วนตัวของธนาคารในเขตกรุงเทพฯ ราคาเล่มละ 165 บาท พร้อมค่าจัดส่ง เล่มละ 13 บาท **พิเศษ** สมาชิกวารสารของซีเอ็ดสั่งซื้อโดยตรงที่บริษัทฯ ได้ในราคาลด 15% เพิ่มค่าลงทะเบียน 4 บาท ค่าจัดส่งฟรี โดยระบุหมายเลขสมาชิกมาด้วย